LETTER

https://doi.org/10.1038/s41586-018-0384-8

CRISPR–guided DNA polymerases enable diversification of all nucleotides in a tunable window

Shakked O. Halperin[1,2,3], Connor J. Tou[1], Eric B. Wong[1], Cyrus Modavi[1,2], David V. Schaffer[1,3,4,5,6]* & John E. Dueber[1,3,7]*

The capacity to diversify genetic codes advances our ability to understand and engineer biological systems[1,2]. A method for continuously diversifying user-defined regions of a genome would enable forward genetic approaches in systems that are not amenable to efficient homology-directed oligonucleotide integration. It would also facilitate the rapid evolution of biotechnologically useful phenotypes through accelerated and parallelized rounds of mutagenesis and selection, as well as cell-lineage tracking through barcode mutagenesis. Here we present EvolvR, a system that can continuously diversify all nucleotides within a tunable window length at user-defined loci. This is achieved by directly generating mutations using engineered DNA polymerases targeted to loci via CRISPR-guided nickases. We identified nickase and polymerase variants that offer a range of targeted mutation rates that are up to 7,770,000-fold greater than rates seen in wild-type cells, and editing windows with lengths of up to 350 nucleotides. We used EvolvR to identify novel ribosomal mutations that confer resistance to the antibiotic spectinomycin. Our results demonstrate that CRISPR-guided DNA polymerases enable multiplexed and continuous diversification of user-defined genomic loci, which will be useful for a broad range of basic and biotechnological applications.

Natural biological systems evolve astounding functionality by sampling immense genetic diversity under selective pressures. Forward genetics emulates this process to help us understand naturally evolved biological phenomena and to direct the evolution of biotechnologically useful material by applying an artificial selection pressure or screen to libraries of genetic variants. New forward genetic approaches would be enabled by a targeted mutator capable of continuously diversifying all nucleotides within user-defined regions of a genome. However, previous targeted continuous-diversification techniques are confined to either evolving specific loci within particular cells under stringent culture conditions[2,3] or mutating particular types of nucleotides in a narrow, user-defined window[4,5] (Extended Data Fig. 1). Conversely, current techniques capable of diversifying all nucleotides within user-defined loci remain discrete owing to their requirement for efficient integration of oligonucleotide libraries at the target site[6,7]. Therefore, no method currently exists to continuously diversify all nucleotides within user-defined regions of a genome (Extended Data Table 1).

DNA polymerases have the capacity to create all 12 substitutions, as well as deletions[8,9]. These enzymes vary in processivity (average number of nucleotides incorporated after each binding event), fidelity (misincorporation rate) and substitution bias (nucleotide bias during misincorporation). In particular, nick-translating DNA polymerases are able to initiate synthesis from a single-stranded break in double-stranded DNA while displacing the downstream nucleotides, and their flap endonuclease domain subsequently degrades the displaced nucleotides, leaving a ligatable nick. We hypothesized that recruiting an error-prone, nick-translating DNA polymerase with a nicking variant of Cas9[10] (nCas9) could offer an ideal targeted mutagenesis tool that is independent of homology-directed repair, and which we term EvolvR (Fig. 1a). The specificity of the polymerase initiation site created by the nCas9 specifies the start site of the editing window, and the mutagenesis window length, mutation rate and substitution bias are controlled by the processivity, fidelity and misincorporation bias of the polymerase variant, respectively.

In the initial design, nCas9 (*Streptococcus pyogenes* Cas9 containing a D10A mutation) was fused to the N terminus of a fidelity-reduced variant of *Escherichia coli* DNA polymerase I (PolI) harbouring the mutations D424A, I709N and A759R (PolI3M)[3]. A plasmid (pEvolvR) expressing the nCas9–PolI3M and a guide RNA (gRNA) in *E. coli* was tested for its ability to mutate a second plasmid targeted by the gRNA over 24 h of propagation. High-throughput targeted amplicon sequencing revealed that the target plasmid accrued substitutions in an approximately 17-nucleotide window 3′ of the nick site (Fig. 1b), consistent with the established 15–20 nucleotide processivity of PolI[11]. Although the sequencing results are probably under-sampling the total diversity generated, we observed substitutions of all four nucleotide types (Fig. 1c). The presence of low-frequency substitutions 5′ of the nick site may be due to endogenous 3′-to-5′ exonucleases removing a few nucleotides 5′ of the nick prior to the polymerase initiating synthesis. Controls expressing an unfused nCas9 and PolI3M with the on-target guide only yielded one low-frequency substitution, whereas expressing nCas9–PolI3M with an off-target guide, as well as expressing nCas9 alone, did not yield any substitutions at a frequency above our detection threshold.

To sensitively quantify the mutation rate and mutagenesis window length of EvolvR variants, we designed a fluctuation analysis[12]. For this assay, the pEvolvR plasmid was co-transformed into *E. coli* with a plasmid (pTarget) containing the *aadA* spectinomycin resistance gene disabled by a nonsense mutation (Fig. 1d). After 16 h of growth, the cultures were plated on spectinomycin and the mutation rates were determined from the number of resistant colony-forming units (CFUs). As shown in Fig. 1e, fluctuation analysis estimated the mutation rate of wild-type *E. coli* to be approximately 10^{-10} mutations per nucleotide per generation, similar to previously reported values[13]. The global mutation rate (the mutation rate of the untargeted genome in cells expressing EvolvR) was determined by measuring the spectinomycin-resistance reversion rate of cells carrying a gRNA targeting *dbpA*, a fitness-neutral RNA helicase gene in the *E. coli* genome[14], whereas the targeted mutation rate was determined with a gRNA nicking 11 nucleotides 5′ of the nonsense mutation in pTarget. Expressing nCas9–PolI3M markedly increased the mutation rate at the targeted locus 24,500-fold over the wild type while increasing the global mutation rate 120-fold over the wild type (Fig. 1e), a global mutation rate comparable to that of previous targeted mutagenesis techniques in *E. coli*[3,6]. By comparison, expressing nCas9 and PolI3M as separate proteins, PolI3M alone, nCas9 alone or a catalytically inactive Cas9 (dCas9) fused to PolI3M, showed significantly lower targeted mutation rates ($P < 0.0001$;

[1]Department of Bioengineering, University of California, Berkeley, Berkeley, CA, USA. [2]University of California, Berkeley–University of California, San Francisco Graduate Program in Bioengineering, University of California, Berkeley, Berkeley, CA, USA. [3]Innovative Genomics Institute, University of California Berkeley and San Francisco, Berkeley, CA, USA. [4]Department of Chemical and Biomolecular Engineering, University of California Berkeley, Berkeley, CA, USA. [5]Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, CA, USA. [6]Helen Wills Neuroscience Institute, University of California, Berkeley, Berkeley, CA, USA. [7]Biological Systems & Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. *e-mail: schaffer@berkeley.edu; jdueber@berkeley.edu

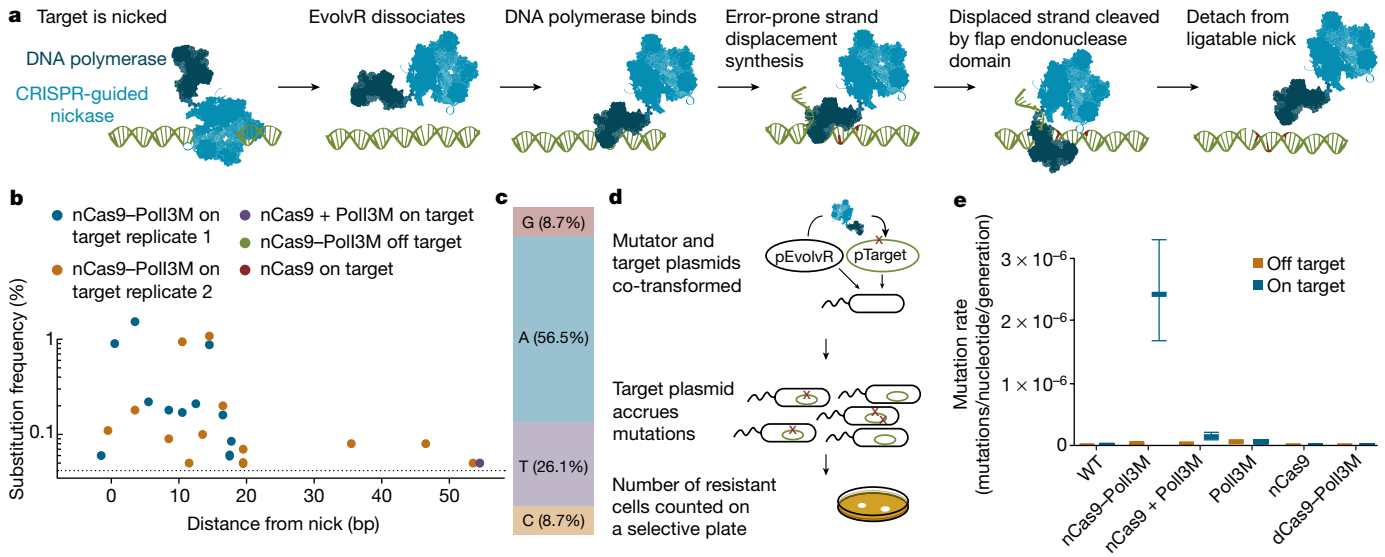


Fig. 1 | EvolvR enables targeted mutagenesis. a, The EvolvR system consists of a CRISPR-guided nickase that nicks the target locus and a fused DNA polymerase that performs error-prone nick translation. **b**, High-throughput sequencing shows that fusing nCas9 to PolI3M resulted in substitutions across an approximately 17-nucleotide window 3′ from the nick. Expressing nCas9–PolI3M with an off-target guide did not show substitutions at a frequency above our detection threshold (dotted line, see Methods 'High-throughput sequencing data analysis'), whereas an unfused nCas9 and PolI3M yielded only one substitution and at low-frequency. **c**, Distribution of the substituted nucleotides; all four nucleotides were substituted by nCas9–PolI3M. **d**, Schematic of the fluctuation analysis workflow used to sensitively quantify targeted and global mutation rates. **e**, The global and targeted mutation rates of wild-type (WT) *E. coli*, nCas9–PolI3M, unfused nCas9 and PolI3M, PolI3M alone, and nCas9 alone were determined by fluctuation analysis. For all figures, 'on target' mutation rates were determined by expressing a gRNA that nicks 11 nucleotides 5′ of the nonsense mutation unless labelled otherwise, whereas the 'off target' mutation rates were determined by expressing a gRNA targeting *dbpA*, a fitness-neutral RNA helicase gene in the *E. coli* genome. Data are the mean of ten biologically independent samples and the error bars indicate 95% confidence intervals. Mutation rates throughout are mutations per nucleotide per generation.

two-sided student's *t*-test). These results suggest that both PolI3M and the nick created by nCas9 are essential for EvolvR-mediated mutagenesis. Expressing nCas9 and PolI3M as separate proteins or PolI3M alone showed a 98-fold or 554-fold increase in global mutation rates compared to wild-type *E. coli*, respectively. Finally, by replacing the D10A nCas9—which nicks the strand complementary to the gRNA—with the H840A nickase, which nicks the strand non-complementary to the gRNA, we found that the direction of EvolvR-mediated mutagenesis relative to the target site of the gRNA is dependent on which strand is nicked (Extended Data Fig. 2).

We hypothesized that the targeted mutation rate could be further increased by promoting the dissociation of nCas9 from DNA after nicking the target locus. Therefore, three mutations (K848A, K1003A, R1060A) that have previously been suggested to lower the non-specific DNA affinity of Cas9[15] were introduced into the fused nCas9. The resulting enhanced nCas9 (enCas9) fused to PolI3M increased the global mutation rate 223-fold compared to wild-type cells (1.9–fold greater than nCas9–PolI3M), yet elevated the mutation rate at the targeted locus by 212,000-fold (8.7-fold greater than nCas9–PolI3M) (Fig. 2a).

PolI3M was initially chosen because it was the most error-prone variant of PolI previously characterized. However, the modularity of EvolvR enables tuning of the mutation rate by using polymerases with different fidelities. First, we confirmed that the fidelity of the polymerase determines mutation rates by comparing enCas9 fused to PolI variants, in decreasing order of fidelity: PolI1M (D424A), PolI2M (D424A, I709N), and PolI3M (D424A, I709N, A759R) (Fig. 2b). Next, to further increase the targeted mutation rate of EvolvR, we screened several additional mutations previously shown to individually decrease wild-type PolI fidelity[3,16,17] (Fig. 3c). Although several of the additional mutations yielded low-activity variants, PolI3M with the additional mutations F742Y and P796H (PolI5M) displayed a mutation rate one nucleotide from the nick that was 7,770,000-fold higher than wild-type cells, and 33-fold higher than PolI3M, making it the most error-prone PolI mutant ever reported. Notably, enCas9–PolI5M did not exhibit either a higher global rate of mutagenesis than enCas9–PolI3M or higher mutation rates than enCas9–PolI3M 11 nucleotides from the nick (Extended Data Fig. 3).

A more processive DNA polymerase could potentially increase the length of the editing window, so PolI5M was exchanged for the more processive bacteriophage Phi29 DNA polymerase (Phi29). Expression of Phi29 variants with previously reported fidelity-reducing and thermostabilizing mutations in combination with the Phi29 single-stranded binding protein showed targeted mutagenesis 56 and 347 nucleotides from the nick site (Extended Data Fig. 4). However, the mutation rate at these distances was not nearly as high as that achieved with PolI3M at shorter distances.

An alternative method to increase the length of the editing window and retain high mutation rates would be to increase the processivity of PolI. Previous work has shown that inserting the thioredoxin-binding domain (TBD) of bacteriophage T7 DNA polymerase into the thumb domain of PolI increases the processivity of the polymerase in the presence of thioredoxin from *E. coli*[18]. Figure 2d shows that whereas the original enCas9–PolI3M did not show a difference between global and targeted mutation rates 56 nucleotides from the nick, incorporation of the TBD into the PolI3M EvolvR gene (enCas9–PolI3M–TBD) produced a 555-fold increase over the global mutation rate at this range. Leveraging this increased editing window length, we targeted enCas9–PolI3M–TBD to a plasmid (pTarget2) containing two nonsense mutations (11 and 37 nucleotides from the nick) in the antibiotic-resistance gene, and thereby showed the ability of EvolvR to generate combinations of multiple mutations with a single gRNA (Fig. 2e).

We hypothesized that unintended translation products consisting of functional DNA polymerase not fused to a functional CRISPR-guided protein contributed to undesirable off-target mutagenesis. Therefore, we codon-optimized the EvolvR coding sequence (enCas9–PolI3M–TBD-CO) to remove three strong internal ribosome binding sites identified using the RBS Calculator[19]. We found that the off-target mutation rate decreased 4.14-fold when expressing enCas9–PolI3M–TBD-CO compared to enCas9–PolI3M–TBD while the on-target mutation rate only decreased 1.23-fold (Extended Data Fig. 5).

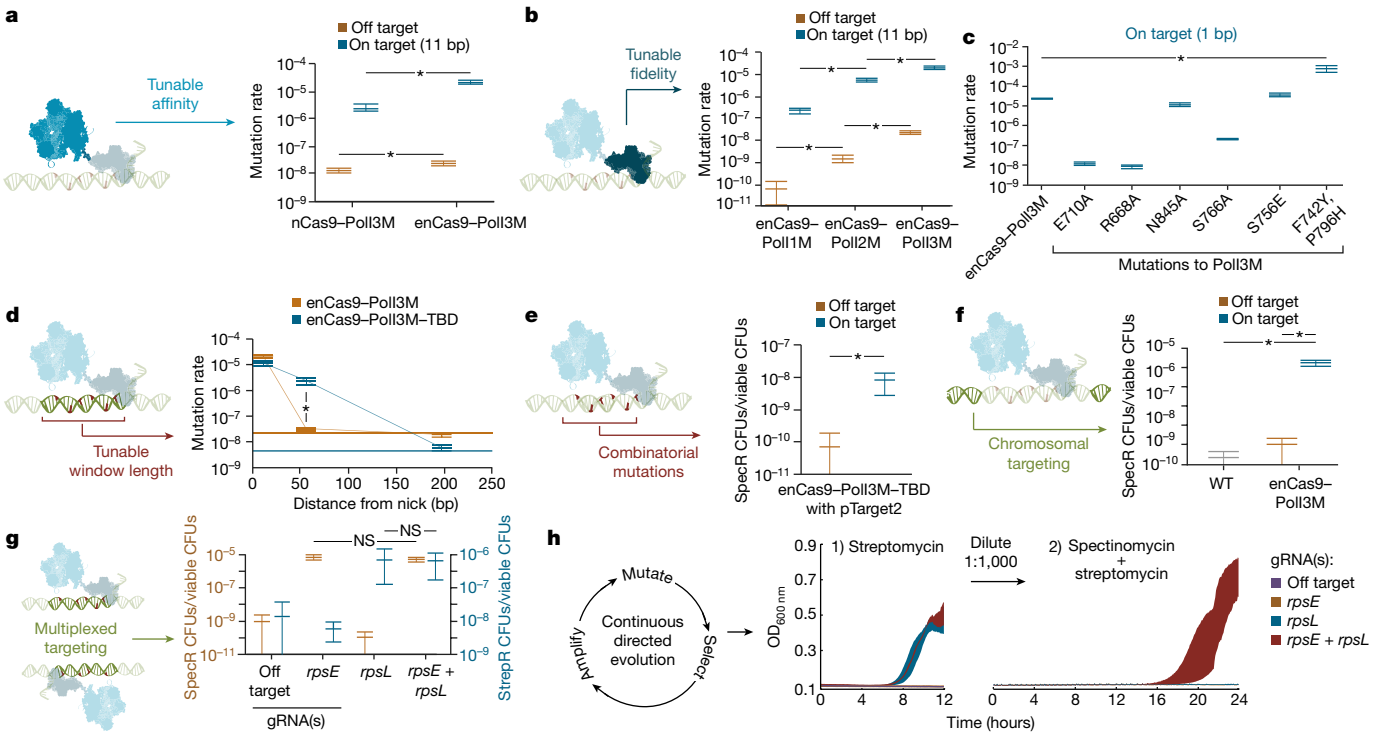


Fig. 2 | EvolvR provides tunable mutation rates and mutagenesis-window lengths, combinatorial mutations, multiplexed targeting and continuous diversification of genomic loci. a, Introducing mutations suggested to lower non-specific DNA affinity into the fused nCas9 (producing enCas9)[15] increased the global mutation rate 223-fold compared to the wild-type mutation rate (enCas9–PolI3M 1.9-fold greater than nCas9–PolI3M), and increased the targeted mutation rate by 212,038-fold over the wild type (enCas9–PolI3M 8.7-fold greater than nCas9–PolI3M). **b**, Mutagenesis rates were dependent on the fidelity of the polymerase. PolI with a D424A mutation (PolI1M) was less mutagenic than PolI with both D424A and I709N mutations (PolI2M), and PolI3M (D424A, I709N, A759R) was the most mutagenic. **c**, Screening mutations in PolI3M previously shown to decrease wild-type PolI fidelity revealed that PolI3M with additional mutations F742Y and P796H (PolI5M) had a mutation rate 7,770,000-fold greater than wild-type cells one nucleotide from the nick. **d**, The editing-window length was increased by incorporating TBD into PolI3M. enCas9–PolI3M–TBD provided a targeted mutation rate 56 nucleotides from the nick that was 555-fold above the global mutation rate, whereas enCas9–PolI3M showed no targeted mutagenesis 56 nucleotides from the nick. **e**, enCas9–PolI3M–TBD targeted to a plasmid containing two nonsense mutations in the spectinomycin resistance gene (pTarget2) showed that EvolvR is able to generate combinations of multiple mutations. **f**, enCas9–PolI3M targeted to *E. coli rpsE* generated approximately 16,000-fold more spectinomycin-resistant CFUs (SpecR CFUs) than when targeted to the *dbpA* locus. **g**, enCas9–PolI3M–TBD targeted to *rpsL* increased the rate of acquiring streptomycin resistance without increasing the rate of acquiring spectinomycin resistance. Coexpression of both *rpsL* and *rpsE* gRNAs increased both spectinomycin and streptomycin resistant CFUs. **h**, Cultures expressing enCas9–PolI3M–TBD and either the *rpsL* gRNA or both *rpsE* and *rpsL* gRNAs grew in streptomycin-supplemented medium, whereas cultures expressing an off-target gRNA or the *rpsE* gRNA did not. After back-dilution into spectinomycin- and streptomycin-supplemented media, only cultures expressing both *rpsE* and *rpsL* gRNAs grew. Mutation rate data are mean ± 95% confidence intervals from ten biologically independent samples. Resistant CFUs/viable CFUs data are mean ± s.d. from ten biologically independent samples. $*P < 0.05$; two-sided student's *t*-test. In **h**, the shaded region of $OD_{600\,nm}$ indicates mean ± s.d. from three biologically independent samples.

The ability to couple EvolvR-mediated mutagenesis to a genetic screen of a non-selectable phenotype would considerably broaden the utility of EvolvR. We found that after targeting EvolvR to a plasmid containing a green fluorescent protein (GFP) cassette with an early termination codon, 0.06% and 0.07% of the population expressed GFP, whereas no cells expressed GFP when an off-target gRNA was used (Extended Data Fig. 6). Importantly, EvolvR also showed the capacity to diversify chromosomal loci by increasing the fraction of the population resistant to spectinomycin approximately 16,000-fold after targeting enCas9–PolI3M to the endogenous ribosomal protein subunit 5 gene of *E. coli* (*rpsE*), which has mutations that are known to confer resistance to spectinomycin[20] (Fig. 2f).

Next, we tested whether EvolvR avoids the toxicity associated with non-targeted mutagenesis systems[21]. We found that, unlike two previously developed non-targeted continuous-mutagenesis systems, EvolvR does not impede cell viability or growth rate (Extended Data Fig. 7a, b). Additionally, targeting EvolvR to the *rpsE* gene evolved more spectinomycin-resistant CFUs per ml compared to these previous non-targeted mutagenesis systems (Extended Data Fig. 7c). Finally, targeting EvolvR to a GFP cassette containing a nonsense mutation resulted in 28 times more GFP-positive cells than when using the most recently developed non-targeted continuous mutagenesis technique (Extended Data Fig. 8).

EvolvR could enable simultaneous diversification of distant genomic loci through coexpression of multiple gRNAs. Expression of a gRNA targeting enCas9–PolI3M–TBD to *rpsL*, a ribosomal protein subunit gene capable of acquiring mutations that confer streptomycin resistance[22], increased the rate of acquiring streptomycin resistance compared to wild-type cells, without altering sensitivity to spectinomycin (Fig. 2g). By comparison, coexpression of the gRNAs targeting *rpsE* and *rpsL* generated approximately the same number of respective spectinomycin- and streptomycin-resistant CFUs as observed for individual expression of the *rpsE* gRNA ($P = 0.0752$; two-sided Student's *t*-test) and *rpsL* gRNA ($P = 0.885$; two-sided Student's *t*-test). This capacity to simultaneously diversify multiple loci will be useful for identifying epistatic interactions. We also note that the expression of two gRNAs that nick separate strands at genomic loci separated by 100 bp was lethal, whereas nicking the same strand at this 100-bp distance was not lethal. Therefore, if multiple gRNAs are to be used to increase the length of the target region, we recommend targeting the same strand.

To evolve resistance to both spectinomycin and streptomycin, we used the continuous diversity generation of EvolvR for continuous

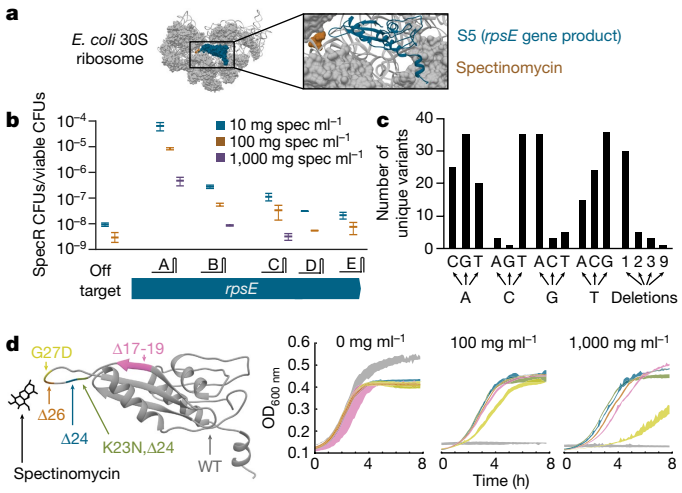


Fig. 3 | EvolvR identified novel mutations to the *E. coli rpsE* gene that confer spectinomycin resistance. a, Spectinomycin inhibits protein synthesis through interactions with the 30S ribosome. **b**, enCas9–PolI3M–TBD targeted to different parts of the endogenous *rpsE* gene with five gRNAs showed higher rates of spectinomycin (spec) resistance than targeting *dbpA* (off-target). Data are mean ± s.d. from three biologically independent samples. **c**, After selection, high-throughput sequencing of the resistant cells containing gRNAs A, B and C revealed that all 12 types of substitutions as well as deletions were generated. **d**, Left, five mutations not previously described as conferring spectinomycin resistance were regenerated in a new strain of *E. coli* (RE1000). Right, growth curves in varying concentrations of spectinomycin confirmed that the mutations provide spectinomycin resistance. Shaded area represents mean ± s.d. from three biologically independent samples.

directed evolution (in which mutagenesis, selection and amplification occur simultaneously) to allow adaptation to modulated selection pressures with minimal researcher intervention. Cultures expressing enCas9–PolI3M–TBD and either the *rpsL* gRNA or both *rpsE* and *rpsL* gRNAs grew in liquid medium supplemented with streptomycin, whereas cultures expressing an off-target gRNA or the *rpsE* gRNA did not (Fig. 2h). After the cultures were diluted 1,000-fold into liquid medium supplemented with both spectinomycin and streptomycin, only cultures expressing both *rpsE* and *rpsL* gRNAs grew.

The clinical utility of spectinomycin as a broad-spectrum antibiotic has motivated previous efforts to characterize genomic mutations conferring spectinomycin resistance[23]. We used the capacity of EvolvR to diversify the genomic *rpsE* gene to identify novel mutations that confer spectinomycin resistance by disrupting the spectinomycin-binding pocket of the 30S ribosome (Fig. 3a). First, we targeted enCas9–PolI3M–TBD to five dispersed loci in the endogenous *rpsE* gene using gRNAs that nick after the 119th, 187th, 320th, 403rd or 492nd base pair within the 504-bp *rpsE* coding sequence (Extended Data Fig. 9a). Then, we challenged the cell populations for growth on agar plates supplemented with varying concentrations of spectinomycin and observed that resistance was highest with the gRNAs targeted to the domain of the ribosomal subunit protein that is proposed to interact with spectinomycin (Fig. 3b). After selection, high-throughput sequencing of the resistant cells containing gRNAs A, B and C revealed that all 12 types of substitutions, as well as deletions, were generated (Fig. 3c). For functional analysis, we introduced five of the candidate mutations not previously described as providing spectinomycin resistance into a different strain of *E. coli* (RE1000) using oligonucleotide-mediated recombination. Growth curves in varying concentrations of spectinomycin confirmed that each of the five mutations (Δ17–19; K23N, Δ24; Δ24; Δ26; G27D) provided varying levels of spectinomycin resistance, but reduced fitness in the absence of selection (Fig. 3d). On the basis of these mutations, we hypothesized that mutations that move Lys26 relative to the spectinomycin-binding pocket confer resistance to spectinomycin by removing a hydrogen bond that stabilizes the interaction of spectinomycin with the ribosome. Therefore we tested an array of deletions that we predicted would move Lys26 and discovered additional novel mutations that confer spectinomycin resistance (Extended Data Fig. 9b, c). This rapid method for discovering genotypes conferring antibiotic resistance will be generally useful for improving the effective use of antibiotics.

EvolvR offers the first example of continuous targeted diversification of all nucleotides at user-defined loci, which will be useful for evolving protein structure and function, mapping protein–protein and protein–drug interactions, investigating the non-coding genome, engineering industrially relevant microbes and tracking the lineage of cell populations that cannot tolerate double-stranded breaks[24]. As a guiding principle for using this tool, our data suggest that 1 μl saturated *E. coli* culture expressing enCas9–PolI3M–TBD for 16 h contains all single substitutions in the 60-nucleotide window with more than tenfold coverage. Future work towards adapting EvolvR for use in cells possessing low transformation efficiency, as well as increasing the mutation rate and window length of EvolvR mutagenesis, would enable new forward genetic applications.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at https://doi.org/10.1038/s41586-018-0384-8.

Received: 1 December 2017; Accepted: 19 June 2018;
Published online 1 August 2018.

1. Ravikumar, A., Arzumanyan, G. A., Obadi, M. K. A., Javanpour, A. A. & Liu, C. C. Scalable continuous evolution of genes at mutation rates above genomic error thresholds. Preprint at https://www.biorxiv.org/content/early/2018/05/03/313338 (2018).
2. Esvelt, K. M., Carlson, J. C. & Liu, D. R. A system for the continuous directed evolution of biomolecules. *Nature* **472**, 499–503 (2011).
3. Camps, M., Naukkarinen, J., Johnson, B. P. & Loeb, L. A. Targeted gene evolution in *Escherichia coli* using a highly error-prone DNA polymerase I. *Proc. Natl Acad. Sci. USA* **100**, 9727–9732 (2003).
4. Ma, Y. et al. Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells. *Nat. Methods* **13**, 1029–1035 (2016).
5. Hess, G. T. et al. Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells. *Nat. Methods* **13**, 1036–1042 (2016).
6. Wang, H. H. et al. Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460**, 894–898 (2009).
7. Costantino, N. & Court, D. L. Enhanced levels of λ Red-mediated recombinants in mismatch repair mutants. *Proc. Natl Acad. Sci. USA* **100**, 15748–15753 (2003).
8. Troll, C., Alexander, D., Allen, J., Marquette, J. & Camps, M. Mutagenesis and functional selection protocols for directed evolution of proteins in *E. coli*. *J. Vis. Exp.* **49**, e2505 (2011).
9. de Boer, J. G. & Ripley, L. S. An in vitro assay for frameshift mutations: hotspots for deletions of 1 bp by Klenow-fragment polymerase share a consensus DNA sequence. *Genetics* **118**, 181–191 (1988).
10. Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
11. Bambara, R. A., Uyemura, D. & Choi, T. On the processive mechanism of *Escherichia coli* DNA polymerase I. Quantitative assessment of processivity. *J. Biol. Chem.* **253**, 413–423 (1978).
12. Sarkar, S., Ma, W. T. & Sandri, G. H. On fluctuation analysis: a new, simple and efficient method for computing the expected number of mutants. *Genetica* **85**, 173–179 (1992).
13. Drake, J. W. A constant rate of spontaneous mutation in DNA-based microbes. *Proc. Natl Acad. Sci. USA* **88**, 7160–7164 (1991).
14. Jagessar, K. L. & Jain, C. Functional and molecular analysis of *Escherichia coli* strains lacking multiple DEAD-box helicases. *RNA* **16**, 1386–1392 (2010).
15. Slaymaker, I. M. et al. Rationally engineered Cas9 nucleases with improved specificity. *Science* **351**, 84–88 (2016).
16. Minnick, D. T. et al. Side chains that influence fidelity at the polymerase active site of *Escherichia coli* DNA polymerase I (Klenow fragment). *J. Biol. Chem.* **274**, 3067–3075 (1999).
17. Loh, E., Salk, J. J. & Loeb, L. A. Optimization of DNA polymerase mutation rates during bacterial evolution. *Proc. Natl Acad. Sci. USA* **107**, 1154–1159 (2010).
18. Wang, Y. et al. A novel strategy to engineer DNA polymerases for enhanced processivity and improved performance in vitro. *Nucleic Acids Res.* **32**, 1197–1207 (2004).
19. Salis, H. M. The ribosome binding site calculator. *Methods Enzymol.* **498**, 19–42 (2011).

20. Funatsu, G., Schiltz, E. & Wittmann, H. G. Ribosomal proteins. XXVII. Localization of the amino acid exchanges in protein S5 from two *Escherichia coli* mutants resistant to spectinomycin. *Mol. Gen. Genet.* **114**, 106–111 (1972).
21. Zheng, X., Xing, X.-H. & Zhang, C. Targeted mutagenesis: a sniper-like diversity generator in microbial engineering. *Synth. Syst. Biotechnol.* **2**, 75–86 (2017).
22. Timms, A. R., Steingrimsdottir, H., Lehmann, A. R. & Bridges, B. A. Mutant sequences in the rpsL gene of *Escherichia coli* B/r: mechanistic implications for spontaneous and ultraviolet light mutagenesis. *Mol. Gen. Genet.* **232**, 89–96 (1992).
23. Brocklehurst, P. Antibiotics for gonorrhoea in pregnancy. *Cochrane Database of Systematic Reviews* **2**, CD000098 https://doi.org/10.1002/14651858.CD000098 (2002).
24. McKenna, A. et al. Whole-organism lineage tracing by combinatorial and cumulative genome editing. *Science* **353**, aaf7907 (2016).

Acknowledgements We thank S. McDevitt at the University of California, Berkeley Vincent J. Coates Genomics Sequencing Laboratory for assistance with high-throughput sequencing, the Arkin laboratory for supplying *E. coli* strain RE1000, W. DeLoache for helping edit our manuscript, and the Innovative Genomics Institute for funding.

Author contributions S.O.H. conceived of all designs, designed the study, contributed to the execution of all experiments, analysed all of the data and wrote the manuscript; C.J.T. contributed to plasmid construction and assay execution for fluctuation analyses, and spectinomycin-resistance mutation identification; E.B.W. contributed to plasmid construction and assay execution for PolI3M mutant screening, Phi29 screening, multiplexing and spectinomycin-resistance identification; S.O.H., C.M., D.V.S. and J.E.D. contributed to assay design. The manuscript was read, edited and approved by all authors.

Competing interests The Regents of the University of California have filed a provisional patent application (62/662,043 and 62/556,127) related to the technology described in this work to the United States Patent and Trademark Office; S.O.H. is listed as the inventor.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0384-8.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0384-8.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to D.V.S. and J.E.D.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Plasmid construction. All plasmids were constructed using a modular Golden Gate strategy. pEvolvR consisted of EvolvR and gRNA expression cassettes, a pBR322 origin of replication and a kanamycin resistance cassette. pTarget consisted of a p15a origin of replication carrying both a functional trimethoprim resistance cassette for selection and a disabled spectinomycin resistance gene (*aadA*) harbouring a L106X nonsense mutation. pTarget2 is identical to pTarget except that the *aadA* gene now carried both Q98X and L106X mutations. The full plasmid sequences are provided in Supplementary Table 1.

High-throughput sequencing of pTarget sample preparation. A pTarget and pEvolvR plasmid were cotransformed into 50 μl chemically competent TG1 *E. coli* prepared by a TSS/KCM method. Cells were allowed to recover in the TSS/LB solution for 1 h, before 4 μl transformation mix was inoculated into 2 ml LB containing 25 μg/ml kanamycin and 15 μg/ml trimethoprim. The cultures were grown for 24 h at 37 °C while shaking at 750 rpm. A 1.5-ml sample of each culture was miniprepped using a Zippy Plasmid Prep kit (Zymo Research).

The oligonucleotides pTarget-F and pTarget-R were used to amplify the target region in a 20-cycle PCR reaction using 100 ng miniprepped DNA as the template. A second PCR reaction added Illumina sequencing adapters and indices to the previous PCR product over 10 thermocycles. A Qubit fluorimeter was used to quantify the DNA before pooling samples. The sample pool was submitted to the University of California, Berkeley Vincent J. Coates Genomics Sequencing Laboratory for quality control and sequencing. Quality control consisted of fragment analysis (Advanced Analytical) and concentration measurement of the sequenceable fraction by quantitative PCR (Kapa Biosystems). The pooled sample was mixed with Illumina PhiX sequencing control library at 10% molarity, diluted to 14 pM, denatured, and run on an Illumina MiSeq using a 150-bp paired-end read MiSeq Reagent Kit v2. Resulting basecall files where converted into demultiplexed fastq format using Illumina bcl2fastq v.2.17.

High-throughput sequencing data analysis. Perfectly complementary paired reads were filtered, and the five randomized nucleotides, amplification primer sequences, and first and last five nucleotides were trimmed using a custom Python script. Bwa and samtools were used to generate alignment files using the wild-type *aadA* gene sequence as a reference. VarScan2 was used for variant calling with the parameters: min-coverage 1; min-reads2 1; variants 1; min-var.-freq 0.0005; p-value 0.99[25]. The limit of detection was determined by sequencing a culture transformed with an empty vector as a control. The highest frequency variant was 0.04% so all substitutions with a frequency under 0.05% were discarded.

Fluctuation analysis assay. A 50-μl sample of chemically competent TG1 *E. coli* were contransformed with pEvolvR and pTarget or pTarget2. After 1 h of recovery at 37 °C, 4 μl was inoculated into a 1.996 ml LB containing 25 μg/ml kanamycin and 15 μg/ml trimethoprim. After shaking at 37 °C for 16 h, 1 ml and 1 μl culture were plated on separate LB agar plates containing 50 μg/ml spectinomycin. For viable CFU counting, 300 μl of 1:50,000,000 diluted culture was plated on LB agar plates. After 24 h of incubation at 37 °C, spectinomycin-resistant CFUs and viable CFUs were counted. Ten replicates were used for each condition.

Calculation of mutation rate and statistics. The Ma–Sandri–Sarkar Maximum Likelihood Estimator was used to determine mutation rates as it is the most accurate and valid for all mutation rates[12]. Falcor was used to calculate the mutation rates by inputing the viable and resistant CFU counts for the ten replicates[26]. A two-tailed Student's *t*-test was carried out to determine *P* values as previously described[27].

Fluorescence-activated cell sorting of EvolvR libraries. pEvolvR expressing either an on- or off-target gRNA was contransformed with pTarget-GFP* and shaken at 37 °C for 24 h. For each sample, the GFP positive fraction of a million events was sorted with a Cell Sorter SH800 (Sony) using a 488-nm laser and a 525/50-nm emission filter.

Continuous evolution of *E. coli* resistant to both spectinomycin and streptomycin. pEvolvR expressing enCas9–PolI3M–TBD and either the off-target gRNA (targeting *dbpA*), *rpsL* gRNA, *rpsE* gRNA or both *rpsL* and *rpsE* gRNAs was transformed into TG1 *E. coli* as previously described. After recovering for one hour, 4 μl of transformation mix was inoculated into 2 ml of LB supplemented with 25 μg/ml kanamycin and cultures were propagated over 16 h at 37 °C. For each culture 2 μl of culture was re-inoculated into 198 μl of LB supplemented with 50 μg/ml of streptomycin. A Tecan M1000 Pro spectrophotometer was used to measure the optical density of each well over 12 h of growth at 37 °C. Each well was then diluted 1,000-fold into LB supplemented with 50 μg/ml of streptomycin and 25 μg/ml of spectinomycin and the optical density of 200 μl of culture was again measured with a Tecan M1000 Pro spectrophotometer over 24 h of growth at 37 °C. Three biological replicates for each gRNA were characterized.

High-throughput sequencing of spectinomycin resistant *E. coli*. A pEvolvR plasmid expressing enCas9–PolI3M–TBD with *rpsE* gRNA A, B, C, D or E was transformed into chemically competent TG1 *E. coli*. Cells were allowed to recover for 1 h before innoculating 4 μl transformation mix into 1.996 ml LB supplemented with 25 μg/ml kanamycin. The cultures were grown for 16 h at 37 °C while shaking. One millilitre and one microlitre of each culture were plated on separate LB agar plates containing 10, 100, or 1,000 μg/ml spectinomycin. Resistant CFUs were counted in the same manner as the fluctuation assays. The colonies from each plate were then pooled into separate cultures containing 2 ml of LB supplemented with 50 μg/ml spectinomycin and grown for 16 h at 37 °C. Genomic DNA was purified using the Wizard Genomic DNA Purification Kit (Promega). One hundred nanograms of purified genome was then processed and sequenced in the same manner as already described for the sequencing analysis of pTarget, with the one alteration that the oligonucleotides rpsE-F and rpsE-R were used for the first round of PCR.

Oligonucleotide recombination. Re-introduction of *rpsE* mutations was performed using RE1000 *E. coli* (MG1655 λ-Red::bioA/bioB ilvG+ pTet2:gam-bet-exo-dam pN25:tetR dnaG.Q576A lacIQ1 Pcp8-araE ΔaraBAD pConst-araC ΔrecJ ΔxonA) developed for recombineering. Electro-competent cells were prepared fresh from overnight cultures of bacteria. The saturated culture was back-diluted 1:70 into 5 ml LB with 100 ng/μl anhydrous tetracycline and shaken at 37 °C until the optical density reached 0.5. Cultures were then transferred to an ice-water bath and swirled for approximately 30 s before being chilled on ice for 10 min. Chilled cultures were centrifuged at 9,800*g* for 1 min. The supernatant was aspirated and the pellet was resuspended in 1 ml ice-chilled 10% glycerol. Washing with glycerol was repeated twice. The final pellet was resuspended in 70 μl chilled 10% glycerol for each transformation. 1 μg of oligonucleotide was electroporated into the cells. The cells were recovered for 1 h at 37 °C in 1 ml LB and streaked out on LB agar plates containing 50 μg/ml spectinomycin. Successful recombination was verified by Sanger sequencing a PCR amplification of the genomic *rpsE* gene.

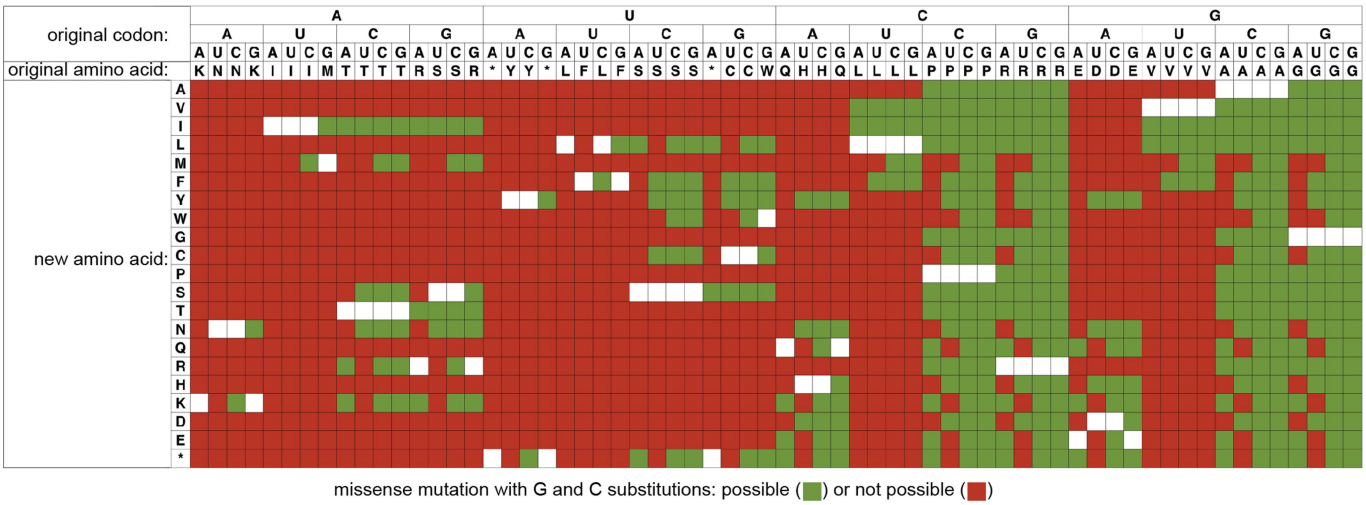
Characterization of spectinomycin resistance. Single colonies of sequence-verified *rpsE* mutants were grown overnight in LB media and then back-diluted 1:200 into LB containing 0, 100 or 1,000 μg/ml spectinomycin. A Tecan M1000 Pro spectrophotometer was used to measure the optical density of each well over 8 h of growth at 37 °C. Three biological replicates of each mutant at each spectinomycin concentration were characterized.

Code availability. The code that support the findings of this study are available at https://github.com/sohhx8/EvolvR.

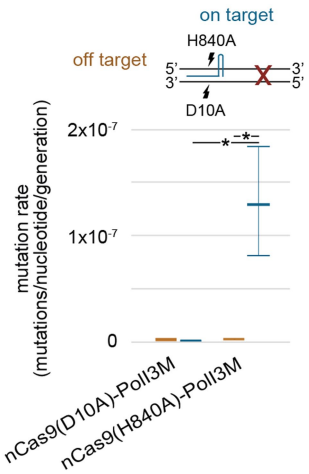
Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

Data availability. The data that support the findings of this study are available from the corresponding authors upon request. High-throughput sequencing data have been deposited as a NCBI BioProject under accession number PRJNA472658. Plasmids encoding enCas9-PolI3M-TBD and enCas9-PolI5M are available from Addgene (plasmids 113077 and 113078).

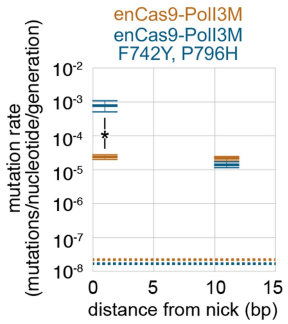
25. Koboldt, D. C. et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* **22**, 568–576 (2012).
26. Hall, B. M., Ma, C.-X., Liang, P. & Singh, K. K. Fluctuation analysis CalculatOR: a web tool for the determination of mutation rate using Luria-Delbruck fluctuation analysis. *Bioinformatics* **25**, 1564–1565 (2009).
27. Rosche, W. A. & Foster, P. L. Determining mutation rates in bacterial populations. *Methods* **20**, 4–17 (2000).
28. Truniger, V., Lázaro, J. M., de Vega, M., Blanco, L. & Salas, M. phi 29 DNA polymerase residue Leu384, highly conserved in motif B of eukaryotic type DNA replicases, is involved in nucleotide insertion fidelity. *J. Biol. Chem.* **278**, 33482–33491 (2003).
29. de Vega, M., Lazaro, J. M., Salas, M. & Blanco, L. Primer-terminus stabilization at the 3′-5′ exonuclease active site of phi29 DNA polymerase. Involvement of two amino acid residues highly conserved in proofreading DNA polymerases. *EMBO J.* **15**, 1182–1192 (1996).
30. Ducani, C., Bernardinelli, G. & Högberg, B. Rolling circle replication requires single-stranded DNA binding protein to avoid termination and production of double-stranded DNA. *Nucleic Acids Res.* **42**, 10596–10604 (2014).
31. Povilaitis, T., Alzbutas, G., Sukackaite, R., Siurkus, J. & Skirgaila, R. *In vitro* evolution of phi29 DNA polymerase using isothermal compartmentalized self replication technique. *Protein Eng. Des. Sel.* **29**, 617–628 (2016).
32. Badran, A. H. & Liu, D. R. Development of potent *in vivo* mutagenesis plasmids with broad mutational spectra. *Nat. Commun.* **6**, 8425 (2015).
33. Greener, A., Callahan, M. & Jerpseth, B. An efficient random mutagenesis technique using an E. coli mutator strain. *Mol. Biotechnol.* **7**, 189–195 (1997).



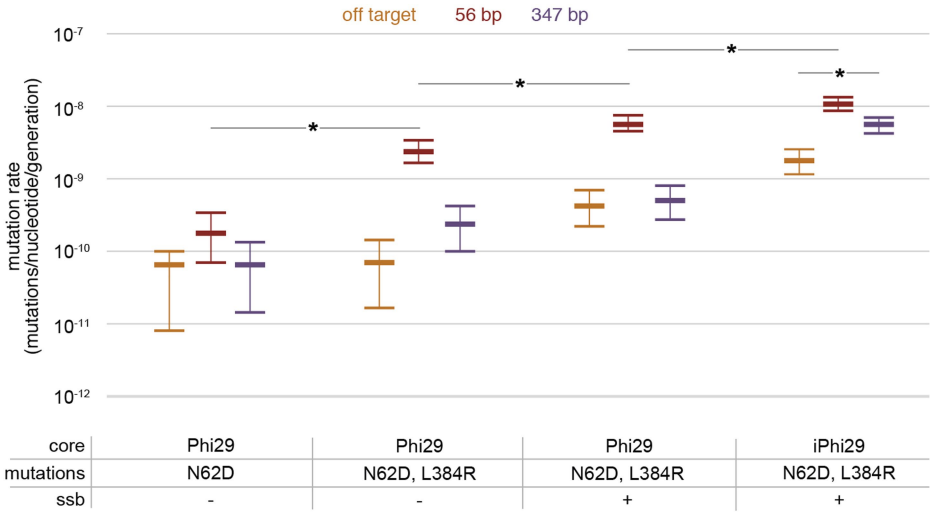
Extended Data Fig. 1 | Bias of cytidine deaminase-mediated targeted diversification. Previous tools enabling diversification of user-defined loci by substituting cytosines and guanines limit the protein coding space that can be explored[4,5]. This chart shows which amino acids can (green) and cannot (red) be reached by mutating cytosines and guanines to any other base for each of the 64 codons, highlighting that only 32% of missense mutations are achievable with targeted cytidine deaminases. The white area depicts the original amino acid identity.



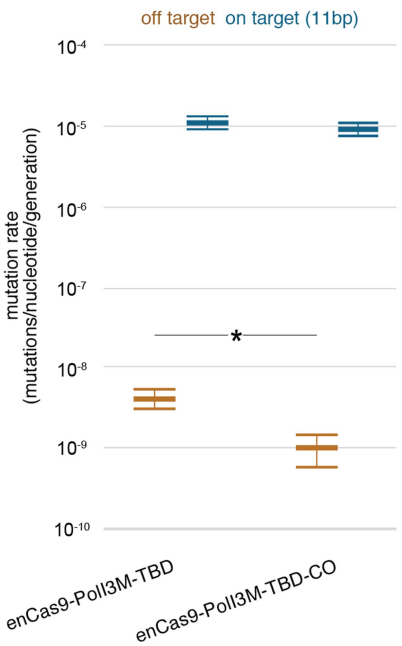
Extended Data Fig. 2 | The direction of EvolvR-mediated mutagenesis relative to the gRNA is dependent on which strand is nicked. Our previous fluctuation analysis in Fig. 1e demonstrated that nCas9(D10A)–PolI3M mutates a window 3′ of the nick site. Here we directly tested whether mutations are generated 5′ of the nick site using a different gRNA. Because DNA polymerases synthesize in the 5′-to-3′ direction, we anticipated that nCas9(D10A)–PolI3M would not provide an elevated mutation rate 5′ of the nick site. We indeed found that expressing a guide RNA which targeted nCas9(D10A)–PolI3M to nick 16 nucleotides 3′ from the nonsense mutation (indicated by a red cross) did not show targeted mutagenesis. We hypothesized that we could induce targeted mutagenesis using the same gRNA by using a Cas9 variant harbouring the H840A mutation, which nicks the DNA strand non-complementary to the gRNA, rather than the D10A mutation, which nicks the strand complementary to the gRNA. nCas9(H840A)–PolI3M increased the mutation rate 16 nucleotides 3′ from the nick by 52-fold compared to the global mutation rate of cells expressing an off-target gRNA. We used the D10A nCas9 variant for all subsequent experiments. Data are mean ± 95% confidence intervals from ten biologically independent samples. *$P < 0.0001$; two-sided Student's *t*-test.



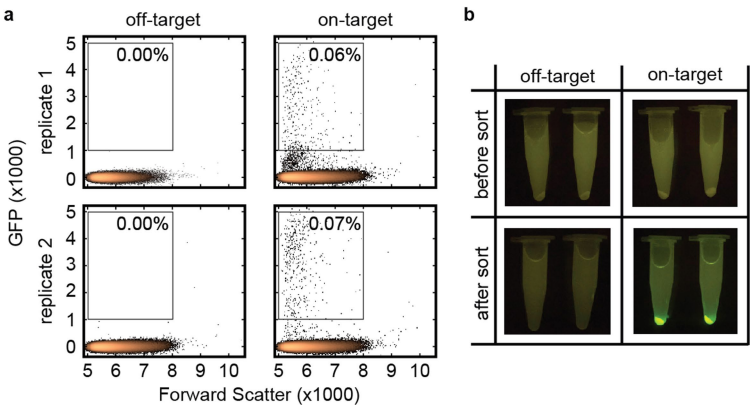
Extended Data Fig. 3 | PolI5M elevates mutation rates 1 nucleotide, but not 11 nucleotides, from the nick compared to PolI3M. PolI3M with additional F742Y and P796H mutations (PolI5M) elevates the mutation rate 33-fold 1 nucleotide from the nick compared to PolI3M. PolI5M did not have a higher mutation rate than PolI3M 11 nucleotides from the nick. Data are mean ± 95% confidence intervals from ten biologically independent samples. *$P < 0.0001$; two-sided Student's *t*-test.



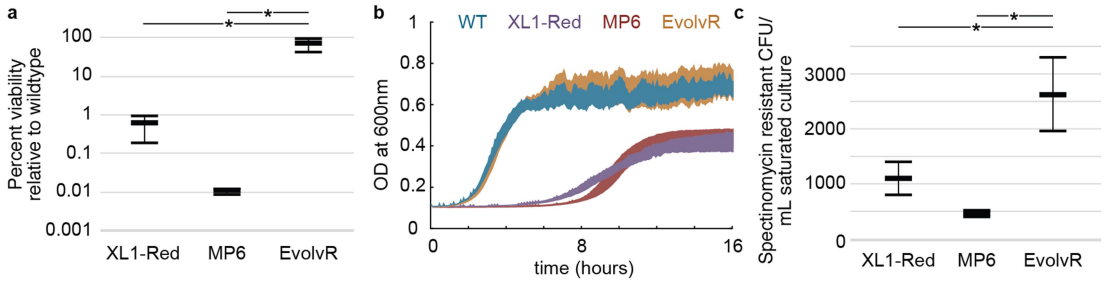
Extended Data Fig. 4 | Fusing a highly processive DNA polymerase to enCas9 increases the target window length. PolI was exchanged for a more processive and higher-fidelity bacteriophage Phi29 DNA polymerase (Phi29). Owing to Phi29 not having a flap endonuclease, residues 1–325 of PolI were inserted between enCas9 and Phi29. Using gRNAs targeting different distances from the nonsense mutation, we found that Phi29 with two previously reported fidelity-reducing mutations (N62D and L384R) elevated the mutation rate 56 nucleotides from the nick compared to the global mutation rate[28,29]. When we expressed Phi29's single-stranded binding protein (ssb), which is known to improve the activity of Phi29, we observed an elevation in the targeted mutation rate[30]. Finally, because the activity of Phi29 is known to decrease at temperatures above 30 °C and the fluctuation analysis was performed at 37 °C, we added mutations previously reported to improve the thermostability of Phi29 (iPhi29) and observed a targeted mutation rate 347 nucleotides from the nick site that was significantly greater than the global mutation rate[31]. Unfortunately, mutations decreasing Phi29's fidelity are known to decrease its processivity explaining our inability to identify Phi29 variants that retain high processivity while offering as high of a mutation rate as PolI3M[28]. Data are mean ± 95% confidence intervals from ten biologically independent samples. *$P < 0.0001$; two-sided Student's *t*-test.



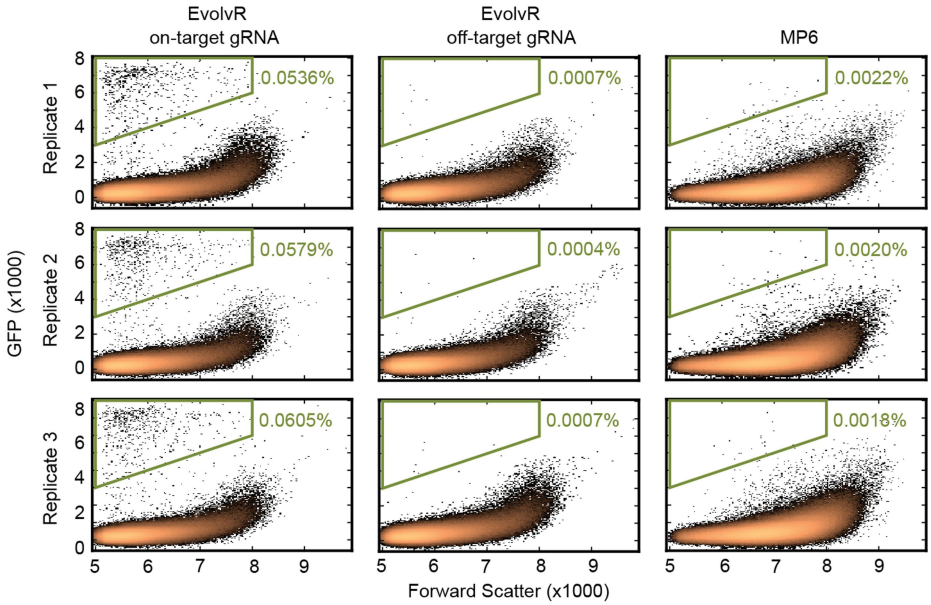
Extended Data Fig. 5 | Removing internal ribosome binding sequences decreases EvolvR-mediated off-target mutagenesis. enCas9–PolI3M–TBD was codon optimized to remove strong ribosome binding sites in the EvolvR coding sequence that were predicted to produce an untethered DNA polymerase. The off-target mutation rate decreased 4.14-fold when expressing enCas9–PolI3M–TBD-CO compared to enCas9–PolI3M–TBD ($P = 0.000482$) whereas the on-target mutation rate only decreased 1.23-fold. Data are mean ± 95% confidence intervals from ten biologically independent samples. *$P < 0.0001$; two-sided student's *t*-test.



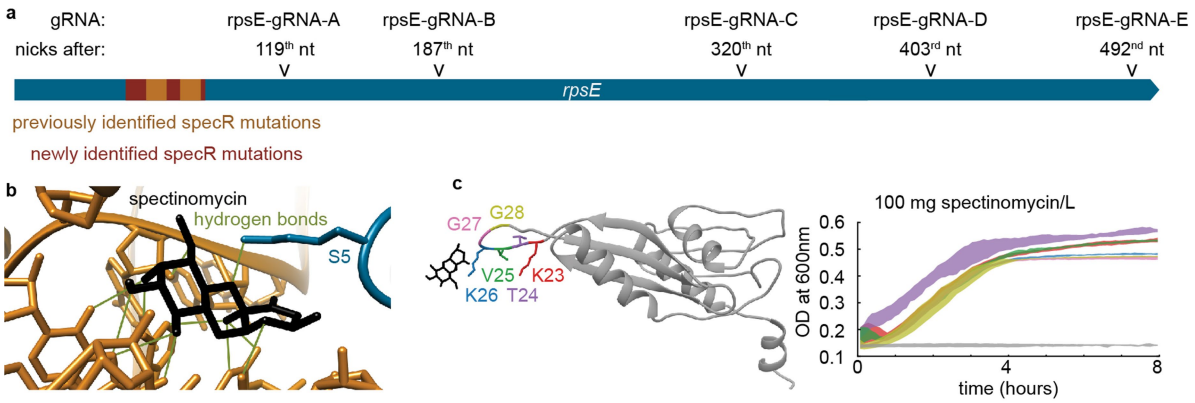
Extended Data Fig. 6 | EvolvR-mediated mutagenesis can be coupled with a non-selectable genetic screen. a, To test the capability for coupling EvolvR-mediated mutagenesis with a non-selectable genetic screen, we designed a target plasmid containing a GFP cassette with an early termination codon in the GFP coding sequence (pTarget-GFP*). After co-transforming pEvolvR with pTarget-GFP* and growing for 24 h, we analysed and sorted the GFP-positive fraction. In the two replicates expressing an off-target gRNA, we did not detect or sort any GFP cells. By contrast, for the two replicates expressing a gRNA nicking four nucleotides away from the chain-terminating mutation in the coding sequence of GFP, we found that 0.06% and 0.07% of the total cells were GFP positive. These results agree with sequencing outcomes from Fig. 1b, which showed that expressing nCas9–PolI3M for 24 h produces substitutions in the target region at frequencies between 0.5% to 1%. **b**, After culturing the sorted populations, both replicates expressing an off-target gRNA did not show growth, whereas both replicates expressing the on-target gRNA grew bright green.



Extended Data Fig. 7 | EvolvR enables targeted genome diversification without affecting viability or growth rate. a, The viability of TG1 *E. coli* expressing EvolvR targeted to the essential *rpsE* gene was significantly higher than TG1 *E. coli* transformed with the MP6 plasmid and induced with 25 mM arabinose and 25 mM glucose (a previously developed plasmid for continuous non-targeted mutagenesis[32], $P = 0.0108$) as well as XL1-Red *E. coli* (a previously developed strain for continuous non-targeted mutagenesis[33], $P = 0.0105$). Viability was measured relative to TG1 *E. coli* transformed with an empty control plasmid. Data are mean ± s.d. from three biologically independent samples. $^*P < 0.05$; two-tailed *t*-test. **b**, TG1 *E. coli* transformed with an empty control plasmid and TG1 *E. coli* transformed with pEvolvR targeting the *rpsE* gene resulted in similar growth curves whereas XL1-Red *E. coli* and TG1 *E. coli* transformed with MP6 plasmid and induced with 25 mM arabinose and 25 mM glucose grew much slower and saturated at lower final optical densities. Shaded area represents mean ± s.d. from three biologically independent samples. **c**, The spectinomycin-resistant CFUs per ml saturated culture of TG1 *E. coli* targeting EvolvR to the *rpsE* gene was significantly higher than XL1-Red *E. coli* ($P = 0.022$) and TG1 *E. coli* transformed with MP6 plasmid and induced with 25 mM arabinose and 25 mM glucose ($P = 0.0049$). Data are mean ± s.d. from three biologically independent samples. $^*P < 0.05$; two-tailed *t*-test.



Extended Data Fig. 8 | EvolvR-mediated mutagenesis performs better than a previous non-targeted diversification technique. To compare the performance of EvolvR and the previously developed non-targeted mutagenesis plasmid MP6 in screen-based directed evolution applications, we co-transformed pEvolvR (enCas9–PolI3M–TBD) or MP6 with a target plasmid containing a GFP cassette with an early termination codon in the GFP coding sequence (pTarget-GFP*). The cultures expressing EvolvR were grown for 24 h and the MP6 cultures followed a two day growth–induction protocol as previously described. Flow cytometry revealed that cultures expressing EvolvR and an on-target gRNA resulted in 28-fold more GFP-positive cells than MP6 cultures.



Extended Data Fig. 9 | Locations of gRNA targets relative to the *rpsE* gene and mutations in ribosomal protein S5 that confer spectinomycin resistance. a, enCas9–PolI3M–TBD was targeted to five dispersed loci in the endogenous *rpsE* gene using gRNAs that nick after the 119th, 187th, 320th, 403rd or 492nd base pair of the 504-bp *rpsE* coding sequence. The locations of the previously identified *rpsE* mutations that provide spectinomycin resistance are coloured orange, and the region where we identified new spectinomycin-resistance mutations is highlighted in red. **b**, The mutations that we discovered confer spectinomycin resistance would be expected to move Lys26 (which is predicted to hydrogen bond with spectinomycin) relative to the spectinomycin-binding pocket. We hypothesized that mutations that move Lys26 relative to the spectinomycin-binding pocket remove that hydrogen bond and destabilize the interaction of spectinomycin with the ribosome, thereby conferring spectinomycin resistance. **c**, Therefore, we tested whether deleting any single amino acid between residues 16 and 35 confers spectinomycin resistance. We found that deleting residues 23, 24, 25, 26, 27 or 28 provides spectinomycin resistance whereas deleting any of the residues between 16 and 22 or 29 and 35 does not. These results support the hypothesis that one mechanism of resistance to spectinomycin is disruption of the interaction between Lys26 and spectinomycin. Data are mean ± s.d. from three biologically independent samples.

Extended Data Table 1 | Comparison of *E. coli* diversification methods

E. coli Diversification Method	Host/culture requirements	Targetability	Ease of use	*In vivo*/ Continuous?	When to use
XL1-Red[33]	Specific Strain	None	Transformation	Yes	Continuous whole genome evolution; target is unknown
MP6[32]	None	None	Transformation	Yes	Continuous whole genome evolution of any strain; target is unknown
Orthogonal polymerase/plasmid[3]	Specific strain and culture requirements	Plasmid	Transformation	Yes	Continuous plasmid evolution; target must be located next to the origin of replication of a specific plasmid
PACE[2]	Specific strain and culture requirements	Phage genome	Custom turbidostat operation, bacteriophage propogation	Yes	Continuous engineered phage genome evolution; target must be inserted within phage genome, and target activity must be coupled to phage propagation
MAGE[6]	Specific strain	User-defined targets	Recombination machinery induction and high-efficiency transformation of oligonucleotide library	No	Generating rationally designed, discrete, user-defined libraries of recombineering strains
EvolvR	None	User-defined targets	Transformation	Yes	Continuous diversification of user-defined genomic loci in any strain

nature research

Corresponding author(s): John Dueber, David Schaffer

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	An online tool named Falcor was used to calculate mutation rates from colony counts.
Data analysis	VarScan2.3.9 was used for variant calling. Python 2.7 was used for high throughput sequencing data analysis. Basecall files where converted into demultiplexed fastq format using Illumina's version 2.17 bcl2fastq.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The data that support the findings of this study are available from the corresponding authors upon reasonable request. (accession codes will be made available before publication)

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences

Study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample size calculation was performed. Sample size for fluctuation analysis was chosen to be 10 because this produced extremely significant differences between groups. Sample size for growth curve analysis was chosen to be 3.
Data exclusions	No data excluded
Replication	All fluctuation analyses were first performed in triplicate. Then we repeated the experiments with n=10. All attempts at replication were successful.
Randomization	Competent cells from the same preparation were randomly distributed between groups for all experiments.
Blinding	A naive researcher counted colonies without knowing experimental conditions for fluctuation analysis

Materials & experimental systems

Policy information about availability of materials

n/a	Involved in the study
☐	☒ Unique materials
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Research animals
☒	☐ Human research participants

Unique materials

Obtaining unique materials	The unique materials (plasmids) that support the findings of this study are available from the corresponding authors upon reasonable request.

Method-specific reporting

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ Magnetic resonance imaging

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☐ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Cells were resuspended in PBS
Instrument	Sony SH800
Software	FlowCytometryTools
Cell population abundance	The purity was determined by plating the sorted fractions and confirming GFP was being expressed\
Gating strategy	We used the bulk cell population

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.